2. Matrix algebra

2.1 Matrix operations

Recall the notation:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \underbrace{\mathbb{R}^{\overset{\#\text{ of rows}}{\overset{\downarrow}{m}} \times \overset{\#\text{ of columns}}{\overset{\downarrow}{n}}}}_{\text{space of matrices}}$$

$a_{ij} \in \mathbb{R}$ is the entry of A that is in the i-th row and in the j-th column.

The columns of A are:

$$A_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix},\ A_2 = \begin{pmatrix} a_{12} \\ \vdots \\ a_{m2} \end{pmatrix}, \ldots,\ A_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \in \mathbb{R}^m$$

Def. For two matrices $A, B \in \mathbb{R}^{m \times n}$, we define the *sum* of A and B, denoted by *$A+B$*, to be the matrix with entries:

$$(A+B)_{ij} := a_{ij} + b_{ij}$$

i.e. addition is defined componentwise. The ij-th entry of A+B is the sum of the ij-th entry of A and the ij-th entry of B

for all $i = 1 \ldots m$, $j = 1 \ldots n$

Similarly, we define *λA* for $\lambda \in \mathbb{R}$ to be the matrix with entries:

$$(\lambda A)_{ij} := \lambda a_{ij}$$

for all $i = 1 \ldots m$, $j = 1 \ldots n$

Ex.
$$\begin{pmatrix} \enclose{circle}{1} & \enclose{circle}{2} & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} \enclose{circle}{0} & \enclose{circle}{1} & 0 \\ 2 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 3 \\ 6 & 7 & 9 \end{pmatrix}$$

→ see "alternative proof" of Lemma 1.13

In consequence:

If $f: \mathbb{R}^n \to \mathbb{R}^m$ linear map given by a matrix $A \in \mathbb{R}^{m \times n}$
and $g: \mathbb{R}^n \to \mathbb{R}^m$ linear map given by a matrix $B \in \mathbb{R}^{m \times n}$
Then $f+g$ is a linear map $\mathbb{R}^n \to \mathbb{R}^m$ given by the matrix $A+B \in \mathbb{R}^{m \times n}$.

Thm. 2.1 (Properties of matrix addition)

(i) $A+B = B+A$

(ii) $(A+B)+C = A+(B+C)$

(iii) $A + \underset{\substack{\uparrow \\ \text{the } m\times n \text{ zero-matrix}}}{0} = A$

(iv) $\lambda(A+B) = \lambda A + \lambda B$

(v) $(\lambda+\mu)A = \lambda A + \mu A$

(vi) $\lambda(\mu A) = (\lambda\mu)A$

Proof: Since sum and scaler multiplication for matrices are defined componentwise, the above properties follow from the according properties for sums and products of real numbers.

Composition of linear fcts. and matrix product

Consider two linear fcts. $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^d \to \mathbb{R}^n$

=> their composition $f \circ g$ is well-defined:

$$(f \circ g)(x) = f(\underbrace{g(\underbrace{x}_{\in \mathbb{R}^d})}_{\in \mathbb{R}^n}) \quad \in \mathbb{R}^m$$

Another scheme to illustrate this:

$$\begin{tikzcd} \mathbb{R}^d \arrow[r, "g"] \arrow[rr, bend right, "f \circ g"'] & \mathbb{R}^n \arrow[r, "f"] & \mathbb{R}^m \end{tikzcd}$$

Exercise: Prove that $f \circ g$ is linear.

(Use the definition of linearity!)

Q: How can we find the matrix for $f \circ g$?

↳ First, what is the size of the matrix of $f \circ g$?

$f \circ g : \mathbb{R}^d \to \mathbb{R}^m$ => the matrix M of $f \circ g$ is in $\mathbb{R}^{m \times d}$

Second, find the actual matrix:

Let A be the matrix for f: $A \in \mathbb{R}^{m \times n}$ and $f(x) = Ax \ \forall x \in \mathbb{R}^n$

Let B be the matrix for g: $B \in \mathbb{R}^{n \times d}$ and $g(x) = Bx \ \forall x \in \mathbb{R}^d$.

Let $x \in \mathbb{R}^d$ (we want M s.t. $f \circ g(x) = Mx \ \forall x \in \mathbb{R}^d$)

$$f\circ g(x) = f(g(x)) = f(Bx) = f(x_1 B_1 + .. + x_d B_d)$$

$$= A(x_1 B_1 + \ldots + x_d B_d) \overset{\text{linearity of } A}{=} A(x_1 B_1) + \ldots + A(x_d B_d)$$

$$\overset{\text{again}}{=} \underbrace{x_1}_{\text{scalar}} \underbrace{AB_1}_{\text{vector in } \mathbb{R}^m} + \ldots + x_d AB_d = Mx$$

$$\text{for } M := (AB_1, \ldots, AB_d) \in \mathbb{R}^{m\times d}$$

This motivates the following definition:

Def. Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times d}$. Then the product of A and B is defined to be the matrix $AB \in \mathbb{R}^{m\times d}$ given by:

$$A\cdot B = (AB_1, \ldots, AB_d)$$

Ex. $A = \begin{pmatrix} 1 & 2 & 5 \\ 0 & 4 & 7 \end{pmatrix}$, $B = \begin{pmatrix} 1 & -1 \\ 0 & -2 \\ 4 & 3 \end{pmatrix}$ $\Rightarrow$ $m = 2$, $n = 3$, $d = 2$

$\Rightarrow AB \in \mathbb{R}^{2\times 2}$

$$\underline{AB_1} = \begin{pmatrix} 1 & 2 & 5 \\ 0 & 4 & 7 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 21 \\ 28 \end{pmatrix}$$

$$AB_2 = \begin{pmatrix} 1 & 2 & 5 \\ 0 & 4 & 7 \end{pmatrix}\begin{pmatrix} -1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 10 \\ 13 \end{pmatrix}$$

$$\Rightarrow AB = \begin{pmatrix} 21 & 10 \\ 28 & 13 \end{pmatrix}$$

Q: What about $A + B$? → not well-defined as A and B not same size.

Q: What about BA? → ok cause $m = d$.

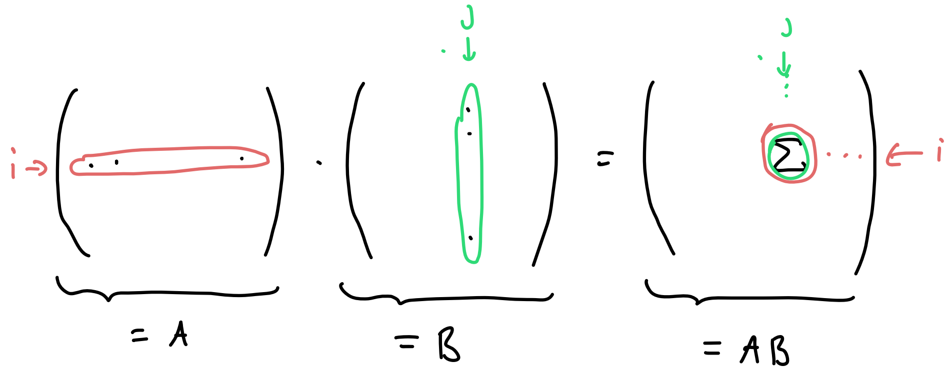
Here is a quicker way to compute matrix products:

Ex

$$AB = \begin{pmatrix} 1 & 2 & 5 \\ 0 & 4 & 7 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & -2 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} 21 & 10 \\ 28 & 13 \end{pmatrix}$$

$(AB)_{ij} = (AB_j)_i$ = i-th row of the vector $AB_j = \sum_{k=1}^{n} a_{ik} b_{kj}$

Schematically :



Notice:
- Matrix multiplication allows us to relatively quickly and easily compute $f \circ g$ for linear fcts. f and g, even if m, n, d are very large.
- To compute $f \circ g(x)$ for all x, it is quickest to compute $A \cdot B$.
- To compute $f \circ g(x)$ for one specific x, it is quickest to compute first Bx then $A(Bx)$.

Thm. 2.2 (Properties of matrix multiplication)

Let $A \in \mathbb{R}^{m \times n}$, B, C matrices, $\lambda \in \mathbb{R}$

(i) $A(BC) = (AB)C$ for $B \in \mathbb{R}^{n \times d}$, $C \in \mathbb{R}^{d \times k}$

(ii) $A(B+C) = AB + AC$ for $B, C \in \mathbb{R}^{n \times d}$

(iii) $(B+C)A = BA + CA$ for $B, C \in \mathbb{R}^{d \times m}$

(iv) $\lambda(AB) = (\lambda A)B = A(\lambda B)$ for $B \in \mathbb{R}^{n \times d}$

(v) $I_m A = A = A I_n$

How to prove these statements?

→ we could check it component-wise using the formula

$$(AB)_{ij} = (AB_j)_i = \sum_{k=1}^{n} a_{ik} b_{kj}$$

→ feasible but rather lengthy, tedious and not very enlightening ☹

→ better idea: use the fact that the matrix product corresponds to composition of linear fcts.

Proof: Let $f(x) = Ax$, $g(x) = Bx$, $h(x) = Cx$

(i) $(A(BC)) \cdot x = f \circ (g \circ h)(x) \mathrel{\textcircled{=}} f(g \circ h(x)) \mathrel{\textcircled{=}} f(g(h(x))) \mathrel{\textcircled{=}} f \circ g(h(x))$

$\mathrel{\textcircled{=}} (f \circ g) \circ h(x) = ((AB)C) \cdot x \quad \forall x \in \mathbb{R}^n$

(The $\textcircled{=}$ hold for all fcts., not just for linear fcts.!)

$\Rightarrow (A(BC))(x) = ((AB)C)(x) \quad \forall x \in \mathbb{R}^n \Rightarrow A(BC) = (AB)C$

Just like in the proof of Thm. 1.12

The proof of (ii) and (iii) is very similar to (i)

(iv) $(\lambda(AB))(x) = (\lambda(f\circ g))(x) = \lambda(f(g(x))$

$= ((\lambda f)\circ g)(x) = ((\lambda A)B)(x)$

similar to (i), (ii), (iii).

$(\lambda(AB))(x) = (\lambda(f\circ g))(x) = \lambda(f(g(x))$

$$\underset{\substack{\uparrow \\ f \text{ linear}}}{=} f(\lambda(g(x))) = (f\circ(\lambda g))(x) = (A(\lambda B))(x)$$

$\Rightarrow (\lambda(AB))(x) = ((\lambda A)B)(x) = (A(\lambda B))(x) \quad \forall x \in \mathbb{R}^n$

$\Rightarrow \lambda(AB) = (\lambda A)B = A(\lambda B)$

(v) Let $h:\mathbb{R}^n \to \mathbb{R}^n$, $h(x) = I_n x = x$

($\Rightarrow$ h is the identity fct. on $\mathbb{R}^n$ $(\forall n\in\mathbb{N})$)

$\Rightarrow (I_m A)(x) = I_m(Ax) = Ax = A(I_n x) \quad \forall x\in\mathbb{R}^n$

$\Rightarrow I_m A = A = A I_n$ □

Matrix algebra vs. algebra of real numbers...

Let A, B be matrices and a, b real numbers.

We will compare whether some usual rules of algebra and arithmetic translate from real numbers to matrices...

- We Know : $a+b = b+a$

  Q : Is it always true that $A+B = B+A$?

  → Yes, by Thm. 1.9.

  Also: $A+B$ is well-defined ⇔ $B+A$ is well-defined. (Size)

- We Know : $ab = ba$ always.

  Q : Is it always true that $AB = BA$?

  → No. This might actually fail in several ways:

  1.) If AB is well-defined, BA does not need to be well-defined. (Size!)

  2.) Even if both are well-defined, the equality might fail (see exercises)

- We Know: If $a+b = a+c$ then $b=c$ (subtract a from both sides)

  Q : Is it always true that: If $A+B = A+C$ then : $B=C$?

  → Yes, because addition of matrices is defined as $\mathbb{R}$-addition in each component.

- We Know : If $a \cdot b = a \cdot c$ (and $a \neq 0$) then : $b=c$ (divide by a)

  Q : Is it always true that: If $AB = AC$ and $A \neq 0$ then $B = C$?

  → No! For example for $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$, $C = \begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix}$:

  $AB = \begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix} = AC$, $A \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $B \neq C$.

- We Know : if $ab = 0$ then : $a=0$ or $b=0$

  Q : if $AB = 0$, then $A = 0$ or $B = 0$?

  → No! For example for $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$:

  $A \cdot B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ but $A, B \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

Matrix-matrix equations

Previously: $\underset{\text{matrix}}{\underset{\nearrow}{A}}\,\underset{\text{vector}}{\underset{\uparrow}{x}} = \underset{\text{vector}}{\underset{\nwarrow}{b}}$ equation $\rightsquigarrow$ want to find vector x

Now: $\underset{\text{matrix}}{\underset{\nearrow}{A}}\,\underset{\text{matrix}}{\underset{\uparrow}{X}} = \underset{\text{matrix}}{\underset{\nwarrow}{C}}$ equation $\rightsquigarrow$ want to find matrix X

Ex: $\begin{pmatrix} 1 & 3 \\ 2 & 0 \\ 0 & 3 \end{pmatrix} \underbrace{\begin{pmatrix} r & s \\ t & u \end{pmatrix}}_{=X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$ Find all solutions X.

In such a small example, we could actually just calculate the product of matrices on the left to obtain six linear equations for the unknowns r, s, t, u and then solve that system.
(You can try it. You'll notice that in this specific example, there is no solution.)

But let us develop a general method for solving matrix-matrix equations...

Given $A \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{m \times d}$, we want to find solutions $X \in \mathbb{R}^{n \times d}$ for the equation: $AX = C$

Write X and C in terms of their colums: $X = (X_1 \ldots X_d)$, $X_i \in \mathbb{R}^n$

$C = (C_1, \ldots, C_d)$, $C_i \in \mathbb{R}^m$

then: $AX = C \iff (AX_1, \dots, AX_d) = (C_1, \dots, C_d)$

$$\iff AX_1 = C_1, \dots, AX_d = C_d$$

Each $\underset{\substack{\nearrow \\ \text{matrix}}}{A}\,\underset{\substack{\uparrow \\ \text{vector}}}{x} = \underset{\substack{\nwarrow \\ \text{vector}}}{C_i}$ is a linear system consisting of m equations and n unknowns

There are d such systems for $C_1 \dots C_d$.

Conclusion: solving $AX = C$ is the same as solving d indep. systems of linear equations $\left(\begin{array}{l}\text{each consisting of } m \text{ equations} \\ \text{and } n \text{ unknowns}\end{array}\right)$

How to solve $AX = C$: A solution X_i of the system $Ax = C_i$ gives the i-th column of the matrix X that solves $AX = C$.

So, we can solve $AX = C$ by solving all the systems $Ax = C_i$ simultaneously by bringing the augmented matrix of $AX = C$

$$(A \mid C) = (A \mid C_1 \dots C_d)$$

into reduced echelon form.

Ex

$$\overset{=A}{\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & -3 & 0 \end{pmatrix}} \cdot \overset{=X}{\underset{X_1 \qquad X_2}{\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{pmatrix}}} = \overset{=C}{\begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}}$$

$\Rightarrow$ augmented matrix

$$(A|C) = \left(\begin{array}{cccc|cc} 1 & 2 & 3 & 4 & 1 & -1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & -1 & -3 & 0 & 1 & 0 \end{array}\right)$$

$$\sim \;\dots\; \sim \begin{array}{c} \\ \left(\begin{array}{cccc|cc} & & & & & c_2 \\ 1 & 0 & 0 & 3 & 1 & -2 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & -1/3 & -2/3 & -1/3 \\ & & & & c_1 & \end{array}\right) \end{array}$$

augmented matrix in echelon form

The equation $Ax = C_1$ has solutions :

$$X_1 = \begin{pmatrix} 1 \\ 1 \\ -2/3 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ -1 \\ 1/3 \\ 1 \end{pmatrix} \quad , \; t \in \mathbb{R}$$

The equation $Ax = C_2$ has solutions :

$$X_2 = \begin{pmatrix} -2 \\ 1 \\ -1/3 \\ 0 \end{pmatrix} + s \begin{pmatrix} -3 \\ -1 \\ 1/3 \\ 1 \end{pmatrix} \quad , \; s \in \mathbb{R}$$

The equation $\boxed{AX = C}$ has solutions :

$$X = (X_1, X_2) = \begin{pmatrix} 1-3t & -2-3s \\ 1-t & 1-s \\ -2/3 + 1/3\,t & -1/3 + 1/3\,s \\ 0+t & 0+s \end{pmatrix} , \quad t, s \in \mathbb{R}$$

$$\Rightarrow S = \left\{ \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ -2/3 & -1/3 \\ 0 & 0 \end{pmatrix} + t \begin{pmatrix} -3 & 0 \\ -1 & 0 \\ 1/3 & 0 \\ 1 & 0 \end{pmatrix} + s \begin{pmatrix} 0 & -3 \\ 0 & -1 \\ 0 & 1/3 \\ 0 & 1 \end{pmatrix} : t, s \in \mathbb{R} \right\}$$

Solution space in parametric form.

Self-products of (diagonal) matrices

Def We call a matrix $A \in \mathbb{R}^{n \times n}$ $(m=n)$ a square matrix.

Def. For $k \in \mathbb{N}$, we define the k-fold self-product of a square matrix $A \in \mathbb{R}^{n \times n}$ by $A^k := \underbrace{A \cdot \ldots \cdot A}_{\#\text{factors} = k}$

And we define $A^0 := I_n$

Ex.

- $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$, $A^2 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 8 \\ 0 & 9 \end{pmatrix}$

  $A^3 = A A^2 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 8 \\ 0 & 9 \end{pmatrix} = \ldots$

- $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $A^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow A^k = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \forall k \geq 2$

- $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{pmatrix}$, $A^2 = \begin{pmatrix} 1^2 & 0 & 0 \\ 0 & 2^2 & 0 \\ 0 & 0 & (-3)^2 \end{pmatrix}$, $A^{\cancel{3}k} = \begin{pmatrix} 1^{\cancel{3}k} & 0 & 0 \\ 0 & 2^{\cancel{3}k} & 0 \\ 0 & 0 & (-3)^{\cancel{3}k} \end{pmatrix} \quad \forall k \in \mathbb{N}$

Def. A matrix $A \in \mathbb{R}^{n \times n}$ is called a diagonal matrix if entries $a_{ij} = 0$ for $i \neq j$.

i.e. $A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}$

$A \in \mathbb{R}^{n \times n}$ is called an upper (resp. lower) triangular matrix if entries $a_{ij} = 0$ for $i > j$ (resp. $i < j$)

i.e. $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1\,n} \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}$ resp. $A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \cdots & a_{n\,n-1} & a_{nn} \end{pmatrix}$

Lemma 2.3

(i) $A, B \in \mathbb{R}^{n \times n}$ diagonal $\Rightarrow AB \in \mathbb{R}^{n \times n}$ diagonal with $(AB)_{ii} = a_{ii} b_{ii}$

In particular, $A^k \in \mathbb{R}^{n \times n}$ is diagonal with $(A^k)_{ii} = a_{ii}^k$

(ii) $A, B \in \mathbb{R}^{n \times n}$ upper (resp. lower) triangular

$\Rightarrow AB \in \mathbb{R}^{n \times n}$ upper (resp. lower) triangular

In particular, $A^k \in \mathbb{R}^{n \times n}$ is upper (resp. lower) triangular

We skip the formal proof as the statements are very obvious and the proof is lengthy and not very enlightening.

The transpose of a matrix

Def. The transpose of a matrix $A \in \mathbb{R}^{m \times n}$ is the matrix $A^T \in \mathbb{R}^{n \times m}$ whose entries are $(A^T)_{ij} := a_{ji}$

So, taking the transpose of a matrix means, we swap the roles of i and j, i.e., rows become columns and vice versa

Ex.

- $A = \begin{pmatrix} 1 & 2 & \boxed{3} \\ 1 & \pi & e \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 1 & 1 \\ 2 & \pi \\ \boxed{3} & e \end{pmatrix}$

  $i=1, j=3$

  i.e

  $i=3$
  $j=1$

- $A = \begin{pmatrix} 2 & 3 \\ -4 & -1 \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 2 & -4 \\ 3 & -1 \end{pmatrix}$

  Observation: For square matrices, the diagonal remains the same

- $A = \begin{pmatrix} 1 & a & b \\ a & 2 & c \\ b & c & 3 \end{pmatrix} \Rightarrow A = A^T$

  We call such matrices symmetric.

  $a_{ij} = a_{ji} \ \forall i,j$

Thm 2.4 $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{n \times d}$

(i) $(A^T)^T = A$

(ii) $(A+B)^T = A^T + B^T$

iii) $(\lambda A)^T = \lambda A^T$

(iv) $(AC)^T = C^T A^T$

(order changed!)

Proof (i) $((A^T)^T)_{ij} = (A^T)_{ji} = A_{ij} \quad \forall i,j$

$$\Rightarrow (A^T)^T = A$$

(ii) $((A+B)^T)_{ij} = (A+B)_{ji} = A_{ji} + B_{ji}$

$$= (A^T)_{ij} + (B^T)_{ij} = (A^T + B^T)_{ij} \quad \forall i,j$$

$$\Rightarrow (A+B)^T = A^T + B^T$$

(iii) very similar to (ii). Try it yourself.

(iv) $$((AC)^T)_{ij} \overset{\text{Def. of T}}{=} (AC)_{ji} \overset{\text{formula for coeff. of matrix product}}{=} \sum_{k=1}^{n} a_{jk} c_{ki} \underset{\text{Def. of T}}{=} \sum_{k=1}^{n} (A^T)_{kj} (C^T)_{ik}$$

$$\underset{\text{multiplication of real numbers is commutative}}{=} \sum_{k=1}^{n} (C^T)_{ik} (A^T)_{kj} \underset{\text{formula for coeff. of matrix product}}{=} C^T A^T \qquad \square$$

As a consequence of (iv), we also get:

$$\underbrace{(A_1 \cdot \ldots \cdot A_k)}_{\text{product of k square matrices}}^T = A_k^T \cdot \ldots \cdot A_1^T$$